59th International Scientific Conference on Information, Communication and Energy Systems and Technologies,
Sozopol, Bulgaria, July 1 – 3, 2024

Performance analysis of cloud service-based data protection systems

Saso Nikolovski[1], Pece Mitrevski[2] and Anita Petreska[3]

Abstract – **This paper identifies a performance evaluation framework with a number of parameters, according to which a solution can be chosen for protection and maintenance of data and information systems in organizational structures. The benefits of the study refer primarily to determining the approach and the stages in the selection of an appropriate solution for protection and recovery of data and information systems, based on their criticality. The chosen concept can be adapted to the operational environment, according to the outlined timeframes set in the business continuity plan, as well as in the plan for recovery from an outage.**

With the conducted analyzes and the conclusions drawn, direct contribution is made to the correct thinking for decision-making when choosing concepts for performance of recovery systems after an outage or catastrophic interruption.

Keywords – **Disaster recovery, Business continuity, cloud service, data protection, data recovery.**

I. Introduction

With modern work processes that are based on digital systems, zero downtime during operational disruption considered from the perspective of sustained business, is an ideal outcome for all organizations. Such an expectation is not always possible or realistically achievable for numerous reasons (outages due to weather events, cyber attacks, etc.) despite the availability of numerous solutions for protection and recovery both in the local data centers of the organizations, as well as solutions based on using a system in cloud [1]. Therefore, from the aspect of management of such organizations, more and more attention is paid to reducing the impact of outages on the overall work process through the maximum estimated outage time that the organization itself can afford, without having permanent consequences for its further operation. When analyzing the outages and setting the goals for consistent recovery from them, while making the same acceptable for the organization, it can be noted that the whole process is based primarily on the time in which the organization is out of operation. and at the same time it includes two periods of time dependent elements. Because of that, using data protection systems to protect and recover data and information infrastructure components, is imposed as a necessary solution for overcoming such critical events to keep business consistency of organizations. In recent times, for such needs, solutions that include in their concept a cloud solution in a partial or full constellation are becoming more and more relevant. Therefore, in the study, a research was done based on a comparative analysis of the performance of two systems - a physical device with a data protection level in the cloud and a solution based entirely on a cloud service.

A. Objectives of the research

By comparing the performance of two different concepts of data protection systems that using cloud services, the research revealed the need to obtain a clear representation of the performability of the solutions for obtaining satisfying time frames when repairing data and systems in organizational structures. Obtaining such time framework will enable the ranking of these systems and their aplicability in relation to the criticality of data and information systems in organizations to maintain their business operability.

II. Related works

The research included in the text of this study is based on a series of papers resulting from the research work in the past ten years, a period in which cloud services became current in the daily operations of company entities [2][3][4][5][6][7][8][9]. In the interest of these researches, several target parameters are constantly circulating, of which recovery point objective (RPO) and data recovery time objective (RTO) are commonly observed, as parameters directly dependent from the performance of data protection systems and their reliability [10]. What is significant to point out is that most of these studies were made in simulation conditions with the most common application of values of the key parameters obtained in an isolated environment without influence from the other infrastructure components. Therefore, the authors themselves point out that the obtained results need their validation in a real production environment.

III. Methodology

Motivated by way of setting theses in the most of mentioned researches, the research conducted in our study is set in a real production environment, to obtain real values for the performance of the considered systems. For this purpose, a server system as virtual machine (VM) has been set up in a real production data center in which software agents of the two data

[1] Saso Nikolovski is with the Faculty of Information and Communication Technologies – Bitola, North Macedonia, E-mail: sasnik@gmail.com

[2] Pece Mitrevski is with the Faculty of Information and Communication Technologies – Bitola, North Macedonia, E-mail: pece.mitrevski@uklo.edu.mk

[3] Anita Petreska is with the Faculty of Information and Communication Technologies – Bitola, North Macedonia, E-mail: anita.petreska1@gmail.com

979-8-3503-8619-6/24/$31.00 ©2024 IEEE

protection systems are placed, with which the backup and recovery operations are performed. In this setting of the data protection systems we apply a reverse engineering approach with which from the agents we take the values for the time and data components obtained during the backup and recovery operations (the time for which the backup or recovery operation was performed and the amount of data that was subject for protection or recovery) in time frame of one year. That means that in this approach, the taken values from agents are used to determine values of derived variables for average data throughput and for the average recovery time for unit of data (MB). On the end, we have used this variables to predict posible behavior of this concepts in scenarios with extreme values of data as a subject of backup and recovery needs, and to determine their applicability (acording time frames) by the criticality of the information systems and their data in the plans for data protection and recovery.

A. Research environment

Considering that the research is conducted within an operational environment critical for the stability of all its systems, a dedicated server system has been established as a test virtual machine.

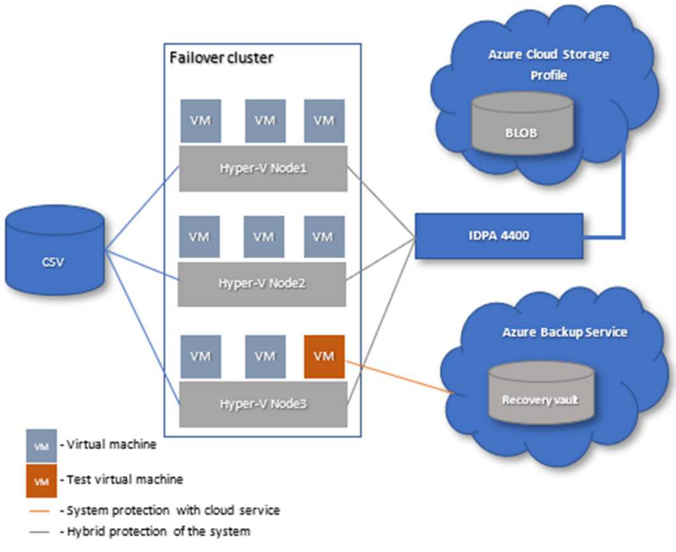


Fig. 1 Production data center – block diagram

As shown in Fig. 1, the production environment in which the research was conducted is composed of three physical server systems (server nodes), placed in a failover cluster with a common storage system and a shared virtual storage volume (CSV). All server systems, including the test system we used for the research, are placed as virtual in the data center with a high degree of redundancy in case of failure of any of the physical components in it [11].

Network connections within the data center (local area network-LAN) are at a speed of 10Gbps, with a symmetrical connection to Internet of 200/200Mbps. In terms of data protection systems, two systems have been deployed, one of which is a Dell EMC IDPA4400 physical appliance [12][13][14] placed in the data center with a storage tier set up in the Azure cloud as a hybrid solution, and the other is Microsoft Azure Recovery Service (MARS), a fully cloud-based system [15].

Software agents from this data protection solutions were instaled in VM to perform backup operations and restore the data in it.

B. Input/output data in modeling aproach

For research purposes, in our study we have create two model types for every used data protection solution. One model was created as basic, where we have used data taken from the agents as inputs used to determine values of derivied variables as basic model outputs. In extended model, test amount of data and derivied variables from basic model are used as inputs for calculation of parameters on outputs, for validation of derivied values taken from the basic model as shown in Fig. 2.

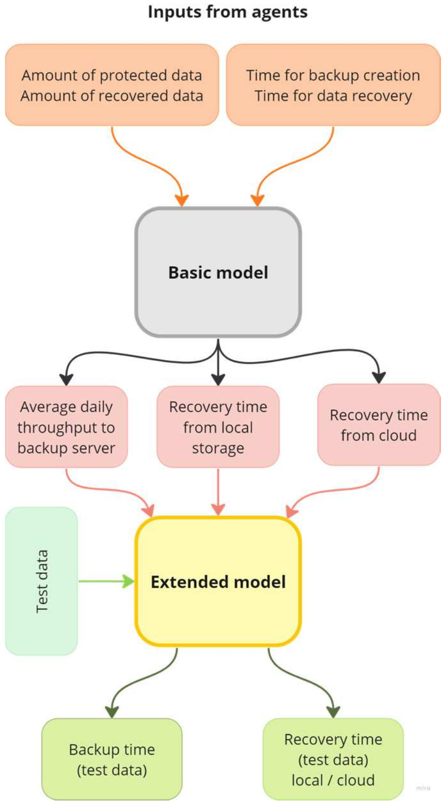


Fig. 2 Input-output parameters in the models

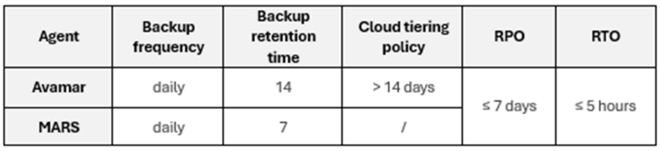
When setting up the models, as shown on Fig. 2, the main emphasis in them is placed on the time to perform backup operation and time to recover protected data.

As a starting point according to which the parameters taken when setting up the models are determined is the business impact analysis of outages (BIA), implemented within the protection policies in the data center.

IV. Simulations and Results

Values that are set as targets in BIA are given in Table I, where the values are separated according to certain parameters for each of the systems separately.

TABLE I
PARAMETERS VALUES IN BIA



Agent	Backup frequency	Backup retention time	Cloud tiering policy	RPO	RTO
Avamar	daily	14	> 14 days	≤ 7 days	≤ 5 hours
MARS	daily	7	/		

In the table, for the hybrid system, the Avamar agent is listed as an agent installed in the server system that is being backed up, and MARS agent for the cloud system.

A. Backup and recovery operations

The processes for creating backup copies, as well as those for recovering data from backup storages, differ due to the concept on which the two systems are based.

In the hybrid system, the backup operation takes place by storing the backup copy in the local storage of the physical device, under the control of set security policies and policies for their retention in the storage (retention policy) for a set period of time (14 days), after which it is transfers to a storage level located in the cloud (cloud tier). Due to this arrangement of the backups in the storage tiers of the system, the recovery of the backups can be from a local storage (if the RPO of the data is within the time frame set in the rention policy) or from the cloud if the data backup due to the rention policy is located in the cloud tier. Therefore, in the presentation of the obtained results from the simulation of the extended model for the hybrid system, the values for recovery from local storage and recovery from the cloud are shown as separate entries.

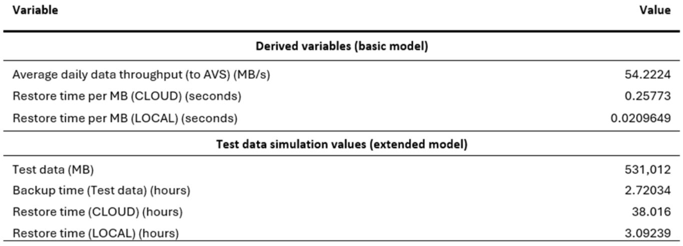
With the MARS system, backup operations take place with two so-called jobs, one of which makes a backup of the complete data structure that is subject to data protection and the second one, to create a system state of the system in which they are used [15].

Due to this concept of backing up the data, during their recovery, this system provides the possibility of granularity when recovering the data, but also the possibility of system state recovery.

B. Simulation results

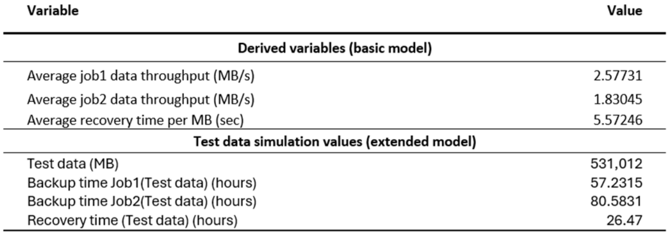
Acording values in Table I, after performing a simulation for the extended models for both systems separately, (basic model is an integral part of the extended model), the output values from the models for the time and data parameters are shown in Table II and Table III.

TABLE II
SIMULATION RESULTS OF A HYBRID SYSTEM



Variable	Value
Derived variables (basic model)	
Average daily data throughput (to AVS) (MB/s)	54.2224
Restore time per MB (CLOUD) (seconds)	0.25773
Restore time per MB (LOCAL) (seconds)	0.0209649
Test data simulation values (extended model)	
Test data (MB)	531,012
Backup time (Test data) (hours)	2.72034
Restore time (CLOUD) (hours)	38.016
Restore time (LOCAL) (hours)	3.09239

TABLE III
SIMULATION RESULTS OF A CLOUD-BASED SYSTEM



Variable	Value
Derived variables (basic model)	
Average job1 data throughput (MB/s)	2.57731
Average job2 data throughput (MB/s)	1.83045
Average recovery time per MB (sec)	5.57246
Test data simulation values (extended model)	
Test data (MB)	531,012
Backup time Job1(Test data) (hours)	57.2315
Backup time Job2(Test data) (hours)	80.5831
Recovery time (Test data) (hours)	26.47

V. COMPARATIVE ANALYSIS

The performance analysis of the considered systems is done by comparing the obtained results from the simulations of the models and it is divided into two parts.

In the first part, a comparison is made of the values obtained in relation to the data transfer when creating a backup, and in the second part, comparison of the values for the time components obtained during the execution of data recovery processes from their backup copies.

TABLE IV
COMPARISON OF DATA TRANSFERS WHEN CREATING A BACKUP COPY

SYSTEM	Hybrid (DP 4400)	CLOUD-BASED (Azure recovery service)	
Component	Average daily data throughput (to AVS)	Average job1 data throughput	Average job2 data throughput
Value (MB/s)	54.2224	2.57731	1.83045

According to the given review in Table IV, it is evident that the hybrid system has many times more data transfer per unit time between the Avamar agent and Avamar server (AVS) when performing the process for creating a backup copy. This is primarily due to the communication connection between the two endpoints of the process, where in the hybrid system scenario, the connection is estabished through the local network (10 Gbps), and in the cloud-based system, the starting point is located in the data center (MARS agent), goes through the Internet service provider (ISP) with 200 Mbps flat rate connection and, it ends in the service located in the Azure cloud. Such changes to data transfers inevitably bring performance degradation, and the result is a total bandwidth with small values, which directly affect the duration of the process by which the copies are created.

The comparative presentation of the simulation results of the data recovery processes with the resulting components that are derived from values obtained from these processes in both systems are shown in Table V.

TABLE V
COMPARISON RECOVERY TIME COMPONENTS VALUES FOR 1MB DATA

SYSTEM	Hybrid (DP 4400)		CLOUD-BASED (Azure recovery service)
Component	Restore time per MB (LOCAL)	Restore time per MB (CLOUD)	Average recovery time per MB
Value (sec)	0.02096	0.25773	5.57246

From the results shown in the Table V, it can be seen that with the hybrid system, the process of recovering data from the device's local storage has a much shorter time than the process that requires data recovery from the cloud storage. This is primarily due to the network connections between the components that participate in the data recovery process, but also to the process of the so-called rehydration of the data that is placed in the storage spaces located in the cloud. This process is a time-consuming component due to the complexity of the data reformation process, including their decryption and decompression before they are returned to the original locations where they came from.

In the case of the cloud-based system, the duration of the data recovery process from the recovery service set up in Azure has many times higher values compared to those required by the hybrid system, which limits the possibilities for wide application of this concept, especially in situations where a short downtime is required in the event of an interruption in the operation and functionality of information systems critical to the sustainability of business processes.

VI. Conclusion

Within the framework of this study, we have presented a research that covers two different concepts of data protection and recovery. Through a comparative analysis of obtained values from the production environment and the results of a simulation performed in given conditions, we have determined the performability of these systems and the possibilities of their application in different scenarios for successful data protection according to the established parameters in the BIA. As a starting point for determining the applicability of the systems that were the subject of observation within the framework of the research, was the ranking of information systems made by Snedaker et al. who in [16] ranks them according to criticality for business operations.

According to the results of the analysis, data protection systems that are critical for maintaining business continuity and that require a recovery time of less than 12 hours can be fully performed by implementing hybrid data protection systems. Vital, important and minor systems in the information infrastructure of business entities can be subject for protection with systems completely placed in the cloud, depending on the amount of data that is subject to protection and the allowed time frame of an outage.

When considering finding an optimal solution for data protection, the best results would be achieved by using hybrid systems for quick recovery of critical functionalities of organizations, and cloud systems would be used for long-term storage (archive) of backup copies and protected data.

References

[1] A. Mathew, C. Mai, "STUDY OF VARIOUS DATA RECOVERY AND DATA BACK UP TECHNIQUES IN CLOUD COMPUTING & THEIR COMPARISON", 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology, 2018, Bangalore, India.

[2] A.A. Tamimi, R. Dawood, L. Sadaqa, "Disaster Recovery Techniques in Cloud Computing", IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), 2019, Amman, Jordan..

[3] A.Mishra, V.Sharma, A.Pandey, "Reliability of Cloud Computing Services", IOSR Journal of Engineering, Volume:04, Issue:01, pp.51-60, 2014.

[4] A.Mishra, V.Sharma, A.Pandey, "Reliability, Security and Privacy of Data Storage in Cloud Computing", International Journal of Advanced Research in Computer Science, Volume 5, No.3, 2014.

[5] H.B.Rebah, H.B.Sta, "Disaster Recovery as a Service: A Disaster Recovery Plan in the Cloud for SMEs", Global Summit on Computer & Information Technology, 2016, Sousse, Tunisia.

[6] J.Mendonça, R.Lima, E.Andradey, J.Araujoy, D.S.Kim, "Multiple-criteria Evaluation of Disaster Recovery Strategies Based on Stochastic Models", 16th International Conference on the Design of Reliable Communication Networks (DRCN), IEEE, 2020.

[7] J.Mendonca, R.Lima, E.Queiroz, E.Andrade, D.S.Kim, "Evaluation of a Backup-as-a-Service Environment for Disaster Recovery", IEEE Symposium on Computers and Communications (ISCC), 2019, Barcelona, Spain.

[8] J.Mendonca, R.Lima, R.Matos, J.Ferreira, E.Andrade, "Availability Analysis of a Disaster Recovery Solution Through Stochastic Models and fault enjection experiments",32nd International Conference on Advanced Information Networking and Applications, IEEE, 2018.

[9] S.Nikolovski, P.Mitrevski, "Data protection and recovery performance analysis of cloud-based recovery service." 2023 58th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST). IEEE, 2023.

[10] S. Nikolovski, P. Mitrevski, “On the Requirements for Successful Business Continuity in the Context of Disaster Recovery”, Proc. of the 57th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST), Ohrid, North Macedonia, 2022.

[11] S.C. Joshi and K.M. Sivalingam, "Fault tolerance mechanisms for virtual data center architectures", Springer Science+Business Media, 2014, New York.

[12] C. Bertrand, M. Keane, "Cloud-ready Data Protection with Dell EMC", ESG White Paper, The Enterprise Strategy Group, 2018.

[13] Scalable Data Protection for Microsoft Windows Server Software-Defined Solutions", White Paper H17894, Dell Technologies, USA, 2019.

[14] Data Protection and Management", Participant guide, Dell Technologies Inc, USA, 2021.

[15] B. Chakraborty, Y. Chowdhury, "Introducing Disaster Recovery with Microsoft Azure: Understanding Services and Tools for Implementing a Recovery Solution", Apress, 2020.

[16] S. Snedaker, C. Rima, "Business Continuity and Disaster Recovery Planning for IT Professionals", Second Edition, Elsevier, 2014.